EVALUATION OF DIFFERENT SEMIVARIOGRAM MODELS IN GROUNDWATER QUALITY MAPPING USING GIS-BASED SIMPLE KRIGING AT KHORASAN RAZAVI-IRAN

NG ZONE FHONG

Thesis submitted in fulfilment of the requirements
for the award of the degree of
B. Eng (Hons.) Civil Engineering

Faculty of Civil Engineering and Earth Resources
UNIVERSITI MALAYSIA PAHANG

JUNE 2015

ABSTRACT

A high demand of the water usage at Khorasan Razavi-Iran was supplied from groundwater. The quality of this groundwater was determined by taking samples from 472 of the wells within the study area like SO_4 and pH. This study was about to choose the best fitted semivariogram models to perform the optimum result of the groundwater quality mapping by using point interpolation. In addition, this research also can show about the estimation accuracy versus the lag distance. The GIS-based software, ILWIS 3.4 was used for geostatistical analysis and generation of the groundwater quality map. The pixel size and lag distance used was 100m. An interpolation technique, Simple Kriging (SK), was applied to obtain the spatial distribution of groundwater quality parameters. In the result, the estimated semivariogram values were the best fitted semivariogram models that were Rational Quadratic model for SO_4 and the Exponential model for pH. Evaluation of these semivariogram models and the estimation accuracy of SK methods was by Coefficient of Determination (R^2), Nash-Sutcliffe Model Efficiency (E), and Root Mean Square Deviation (RMSD). For SO_4 in Rational Quadratic model, it was 0.95 for R^2 which represent the highest value among others model, 0.95 for E which represent the highest also and 1.67 for RMSD which indicate the lowest error from the rest model. On the other hand, the predicted error for pH in Exponential model was 0.78 for R^2, 0.77 for E and the last was 0.18 for RMSD. Furthermore, the accuracy trend was illustrated in table form and figures that the error of the accuracy is getting worse for example the R^2 for SO_4 in 500m lag with 0.96 was become lesser in 1500m that is 0.69. For the E from 500m to 1500m, the value was lesser as well as RMSD was increased due to the distance. All in all, the objectives were accepted.

ABSTRAK

Permintaan yang tinggi pada penggunaan air di Khorasan Razavi-Iran adalah dibekalkan dari air bawah tanah. Kualiti air bawah tanah ini telah dikumpulkan dengan mengambil sampel dari 472 buah telaga di dalam kawasan kajian seperti SO_4 dan pH. Objektif kajian ini adalah untuk memilih model semivariogram yang terbaik untuk melaksanakan keputusan yang optimum untuk pemetaan kualiti air bawah tanah dengan menggunakan titik interpolasi. Di samping itu, kajian ini juga boleh menunjukkan tentang ketepatan anggaran berbandingan jarak lag. Perisian berasaskan GIS, ILWIS 3.4 telah digunakan bagi analisis Geostatistik dan penjanaan peta kualiti air bawah tanah. Saiz piksel dan jarak lag digunakan ialah 100m. Teknik interpolasi, 'kriging mudah' (SK), telah digunakan untuk mendapatkan taburan parameter kualiti air bawah tanah. Untuk hasil didapati, nilai semivariogram dianggarkan terbaik dipasang dalam model ialah rasional kuadratik model bagi SO_4 dan model Eksponen merupakan model yang terbaik bagi pH. Penilaian model semivariogram dan ketepatan anggaran dalam kaedah SK adalah dengan Pekali Penentuan (R^2), Kecekapan Nash-Sutcliffe Model (E), dan Root Mean Square Sisihan (RMSD). Untuk SO_4 dalam model kuadratik rasional, ia memberi nilai 0.95 adalah untuk R^2 di mana merupakan nilai tertinggi di kalangan model lain, E adalah 0.95 di mana nilai yang tertinggi juga dan RMSD adalah 1.67 iaitu kesilapan yang paling rendah daripada model yang lain. Seterusnya, kesilapan yang diramalkan untuk pH dalam model Eksponen adalah 0.78 untuk R^2, E adalah 0.77 dan RMSD adalah 0.18. Selain itu, trend ketepatan itu digambarkan dalam bentuk jadual dan angka bahawa kesilapan ketepatan semakin teruk di mana R^2 untuk SO_4 dengan 500m lag ialah 0.96 telah menjadi lebih kecil dengan 1500m lag iaitu 0.69. Bagi E dari 500m hingga 1500m lag, nilai itu telah menjadi semakin kecil di samping RMSD telah meningkat disebabkan oleh jarak meningkat. Oleh itu, semua objektif telah diterima.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GIS	Geographical Information System
WEPA	Water Environment Partnership in Asia
ILWIS	Integrated Land and Water Information System
IDW	Inverse Distance Weighting
OK	Ordinary Kriging
SK	Simple Kriging
AK	Anisotropic Kriging
UK	Universal Kriging
CK	Co-Kriging
GNU	General Public License
JMG	Minerals & Geoscience Department Malaysia

LIST OF SYMBOLS

MLD	Million liter per day
SO_4	Sulphate
pH	Negative log of the activity of the hydrogen ion in an aqueous solution
km^2	Kilometer square
Ca^{2+}	Calcium
Na^+	Sodium
Mg^{2+}	Magnesium
Cl^-	Chloride
HCO_3^-	Bicarbonate
NO_3	Nitrate
TDS	Total dissolved solids
UV	Ultraviolet
ppm	Part per million
mg	Milligram per liter
n	Number of the untried points used to test
y_i	Real values from input
$\hat{y}_i$	Predicted value from the kriging model
N_h	Number of pairs separated by vector h
h	Lag distance
x_i	The starting location
$x_i + h$	The ending location

γ	The only dependent on the length of lag distance but not its direction
R^2	Coefficient of determination
E	Nash–Sutcliffe model efficiency coefficient
RMSD	Root mean square deviation
N	Number of observations used to fit the model
Σ	Summation symbol
x	Mean x value
y_i	Y value for observation i
y	Mean y value
σ_x	Standard deviation of x
σ_y	Standard deviation of y
X_{obs}	Observed values
X_{model}	Modelled values at time/place i
u, u_α	Location vectors for estimation point and one of the neighbouring data points, indexed by α
n(u)	Number of data points in local neighbourhood used for estimation of Z*(u)
m(u), $m(u_\alpha)$	Expected values (means) of Z(u) and $Z(u_\alpha)$
λ_α	Kriging weight assigned to datum $Z(u_\alpha)$ for estimation location u; same datum will receive different weight for different estimation location.

CHAPTER 1

INTRODUCTION

1.0 BACKGROUND OF PROPOSED STUDY

Groundwater is a hidden source of water supply because it remains in the subsurface of earth. To date, groundwater is the vital source of freshwater. The raise on the freshwater demand for irrigation, industrial purpose, drinking water and domestic water supply has become the main factors on the groundwater exploitation. However, groundwater is still the least utilized sources compared to the surface water resources such as lakes, Perennial River and rainfall. There are many potential spots of groundwater where can make use of them in the process of clean water. To be safe in the process of clean water making, the quality of the groundwater is the priority issue before it can be consumed (Ibrahim *et al.*, 2012).

Groundwater is normally high in quality. After being filtered naturally via the different layers of soil properties, a colourless, clear and free from microorganism contaminated water formed, but the groundwater is always maintained in such excellent quality properties. As a result, the groundwater quality assessment is crucial to the nation so that a better quality of clean water for drinking, domestic use and so on.

Geographical Information System (GIS) is a powerful tool to help on the prediction of the natural resources and environmental problems for example the groundwater quality. There are many different applications of GIS in the groundwater studies, like groundwater movement modelling, site suitability analyses, groundwater quality assessment with spatial data for prediction purpose (Nas and Berktay, 2010).

In GIS application, all the spatial data is going to be analyzed by Kriging method, a method of interpolation which helps in estimation and unknown value prediction from data collected at specific study area. In addition, this method has a benefit to minimize the error of predicted values which are estimated by spatial distribution. Kriging uses semivariogram to show the spatial variation. The main problem of Kriging is to choose the best fitted semivariogram model, which affect the spatial interpolation accuracy of the groundwater quality mapping. There are many different types of semivariogram model such as Circular model, Spherical model, Exponential model, Gaussian model, and so on (Huang *et al.*, 2012).

A semivariogram is one of the significant functions to show the spatial correlation in observations measured in sample of the study area. It is represented as a graph that shows the semi variance in measure with distance or average lag between all pairs of sampled locations. By producing this graph, a mathematical model that describes the variability of the measure with location can be built. Modeling of the relationship between sample locations to indicate the variability of the measure with distance of separation is called semivariogram modeling. It is applied to applications involving estimating and the predicted value of a measure around the new location. Semivariogram modeling is also called as semivariogram modeling (Gandhi, 2012).

1.1 PROBLEM STATEMENT

There are too many different types of semivariogram models in kriging interpolation analysis. There are all together six different types of semivariogram models to be tested to determine which the best fitted semivariogram model is based on SO_4 and pH in the generation of groundwater quality map. Besides that, there are also various kind of Kriging method to deal with different situations or problems.

In the year 2010, Malaysia water demand was increased abruptly to approximately 17,000 MLD as our country's population achieves 30 million. There are almost 98% of water consumed from the surface water resources now in Malaysia. Due to the excessive used of water resources on the surface water, it may cause the water resources unsustainable and quick action must be taken to overcome this problem (Darby, 2009).

Surface water resources are directly exposed to the environment, thus they are likely to be affected by the land use factors, individual behavior and also the extreme weather condition. Therefore, groundwater is the best option to replace the in danger resources since it is a sustainable and reliable water source as long as it is extracted in a right way by using appropriate technology and steps. This problem must be taken into consideration for future prospect and to be concerned with, all the respective organizations so that our nation have a better water supply.

1.2 RESEARCH OBJECTIVE

The objectives of this study are:-

- To determine the best fitted semivariogram model to determine spatial variations of groundwater quality based on the SO_4 and pH.

- To determine the trend of error propagation in interpolation of groundwater quality data from the point sources.

1.3 SCOPE OF PROPOSED STUDY

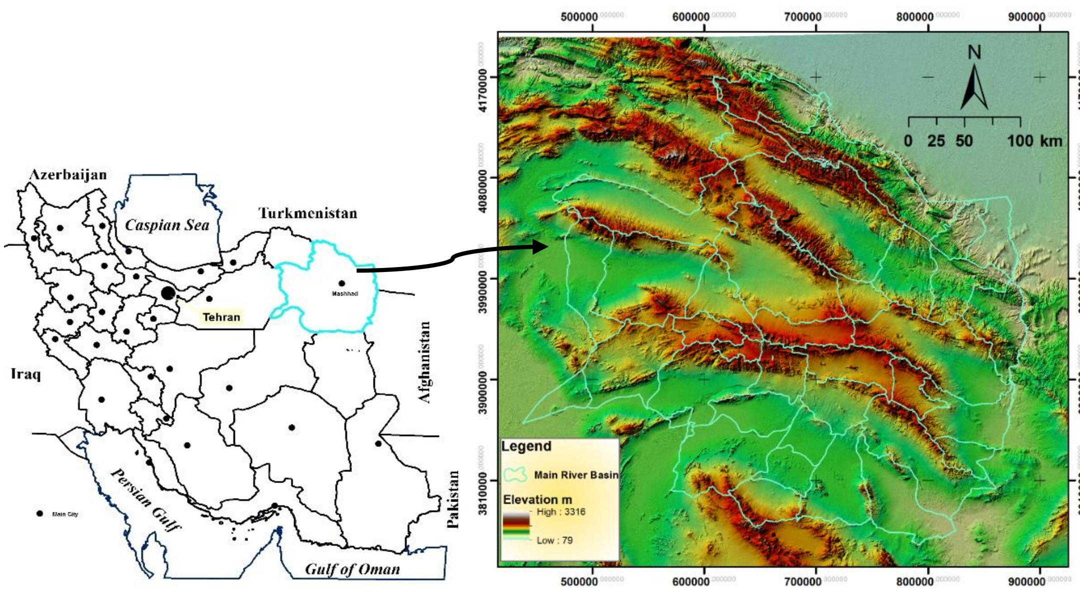

Figure 1.1: *The locality map of the Khorasan Razavi-Iran*

Source: Akbari (2014)

There are 472 wells coordinate and their respective groundwater quality parameters, data which are SO_4 and pH being selected in this research. The groundwater quality parameter is based on the pH value and Sulphate (SO_4) in the year of 2012.

The software for the spatial interpolation is called ILWIS 3.4, a vector and raster processing, GIS-based free download software (Bhd, 2011). It is used for point interpolation by using Simple Kriging. The pixel information and the lag distance used was 100m.

The Khorasan Razavi province is located in the north-eastern part of Iran (Figure 1.1). Approximately, this province is 127432 km^2, and its population is estimated at over 6047000. This city mainly relied on groundwater resources as their daily purpose, for example, irrigation, domestic purpose, drinking and industrial purpose.

1.4 EXPECTED OUTCOME

- To obtain the best fitted semivariogram model in the spatial variation of groundwater quality based on the SO_4 and pH.

- To check the trend of error propagation in interpolation of groundwater quality data from the point sources.

1.5 SIGNIFICANCE OF STUDY

The first importance of this study is to select the best fitted semivariogram model for kriging interpolation which can be evaluated by the estimation error measure such as the Coefficient of determination, R^2; Nash–Sutcliffe model efficiency coefficient, E, and Root-mean-square deviation, RMSD. The best semivariogram models in groundwater quality mapping can be known according to different kind of water quality parameters. It can help the next researcher to choose as a control map for comparison. This research has contributed to the water quality modeller on choosing a suitable semivariogram model when undergoing spatial interpolation by Kriging method.

Next, the second importance of this study is to define the capabilities of GIS techniques in order to perform the groundwater quality assessment in a specific watershed area. It is also the best tools assisting in decision making by utilized GIS. For example, in planning and developing scenarios, water quality zonation provide a guideline for future development and may take a better action in groundwater management. It also helps in finding out a better solution for a quick assessment.

1.6 CONCLUSION

In this chapter, the background of proposed study, problem statement, our research objective, the scope of proposed study, the expected outcome of our project, and also the significance of the study have been discusses. In this study, the different semivariogram

model in groundwater quality mapping by using simple kriging method was the main. A
this study, we will proceed with our chapter 2, Literature Review.

CHAPTER 2

LITERATURE REVIEW

2.0 INTRODUCTION

Groundwater plays a crucial role in drinking water for everyone, especially for those who lives in rural areas. However, it may become dirty and polluted due to the factors of human activities and also natural factors. Different land use such as industrial, agricultural works and many more activities will influence the groundwater quality. These contamination problems may lead to disease, infection, no water supply, economy rise up, and so on (Nas and Berktay, 2010).

For the physical and chemical condition of groundwater, the best solution need to be found to control the groundwater quality for a better future life. Therefore, a good monitoring network should be shown of both adequate and appropriate information concerning the groundwater quality as well as be effective in terms of cost (Narany *et al.*, 2014).

GIS is being applied in this study and kriging techniques for groundwater quality mapping are used such as simple kriging and ordinary kriging. In this study, Kriging techniques were used to interpolate groundwater parameters pH values and SO_4 in the study area. Besides that, there are many different semivariogram models need to be plotted and selected the best fitted model to increase the accuracy of results obtained.

2.1 GROUNDWATER QUALITY PARAMETERS

According to Resources (2013), groundwater quality standards are the maximum allowable concentrations of pollutants in groundwater which may be tolerated without creating a threat to human health or which would otherwise render the groundwater unsuitable for use as a drinking water source.

Groundwater consists of many different chemical compounds with their respective concentrations. The soluble minerals in soils and sedimentary rocks are the main source of chemical constituents in groundwater (Committee, 2005). There are about 95% of the ionic compound in groundwater like the cations potassium K^+, calcium Ca^{2+}, sodium Na^+, magnesium Mg^{2+} as well as the negatively charged chloride Cl^-, sulphate SO_4^{2-}, bicarbonate HCO_3^- and nitrate NO_3^-. The combination of these ions will be same as the total dissolved solids TDS.

The parameters used in this research to determine the quality of the groundwater are pH value, and sulphate SO_4 because they are the fundamental properties in groundwater which were the best water quality indicator among the rest.

2.2 pH VALUES

The pH value of groundwater is a foremost property to declare its alkalinity and acidity. Due to the change of the pH value, it may affect the chemical properties of many organic and inorganic substances dissolved in the groundwater. It is a good indicator for the groundwater usage, for example to determine the suitability for domestic and commercial uses were mainly for drinking purpose. It's also helpful in the balancing of the harmful chemicals through water flow (Baskaran Sundaram and Ross S. Brodie, 2009).

For the chemical properties of the groundwater, it will be dependable to the duration of the groundwater surface contact with the rock or sediment particles. The pH value will be stabilized or act as a buffer by these chemical composition of the bedrock. Therefore, the time is the only factor which can change the groundwater characteristics. As the time goes by, it will make the rock to be mineralized and yet the pH value will be differ in time because of the different mineral content react with the water itself.

There are 3 parts of results for the pH value. Firstly, the groundwater with 7 is the neutral form of water, which we can know it as a distilled water. If the groundwater has a pH value greater than 7, it is a basic water, whereas it is an acidic water as the pH value is less than 7. Groundwater, which is passing through the carbonate rocks like marbles and limestones will ordinarily possess a pH value less than 7 as the acidic water. These groundwater had gone through a process called neutralisation. Normally, the groundwater will remain it status as an acidic water, although there are few carbonate rocks such as volcanic rocks, sandstones, gneisses, metamorphic granitic schists and so on (Agwt.org, 2003).

2.3 SULPHATE, SO_4

Sulfate ions are considered soluble ions which will dissolve in groundwater that give certain concentration in the water. Many of the sulfate ions are formed when there is the ores undergo an oxidation process. Besides that, these ions also can be found in the industrial wastes. Thus, the direct method of sulphate concentration in groundwater was from the diffusion of industrial waste into the water table. An UV Spectrophotometer was used to measure its concentration. The desirable limit for Sulphate is 200 and 400 mg/l in Permissible limit according to IS: 10500-2012 (Dohare *et al.*, 2014).

There are many minerals that consist of sulphate such as sodium sulfate (Glauber's salt), calcium sulfate (gypsum), and magnesium sulfate (Epsom salt). 1 mg/L is equivalent to 1000th of 1 gram in 1 liter of water, which is approximately same as one part per million (ppm). In addition, 1 ppm is one drop of the sulphate in ten gallons of water. As sulphate in water exceeds 250 mg/L, it will taste bitter which is disaggreable for drinking. There are also many other effects if a high level of sulphate concentration like they may corrode the plumbing especially particularly copper piping (Health, 2014).

Sulphate basically possess a maximum concentration of 50 mg/l in natural waters as well as 1000 mg/l for the water where coal, pyrite, and lignite is contacted or inside the water. It can be classified into 3 different stages for example, poor (0-200 mg/l), good (200-400 mg/l), and bad (>400 mg/l) .

2.4 SPATIAL PATTERN

2.4.1 Geo-Statistic

According to Nas (2009), GIS has greatly benefited based on the natural resources and environmental concerns for example groundwater. Geostatistical analysis was a helpful system to calculate the water variables in space and time.

Besides, geostatistical methods were developed to present spatial correlation models with a semivariogram. The semivariogram is usually found in geostatistics and there is an interpolation method called kriging (Cinnirellaa *et al.*, 2005).

Georges Matheron was the first man developed about geostatistics has made rapid development along the years. Now, geostatistics term has been widely used to describe all the statistics application in hydrogeology and geology where the attribute data is stochastic in space. Geostatistical techniques are frequently used to calculate the estimated parameters in deterministic models because the heterogeneity of the subsurface is very hard to evaluate accurately for the use in deterministic models. Due to the groundwater flow problems, attributes like water level and salinity are collected at a few number of sites whereas values at the un-sampled sites are needed to further its analysis. As a conclusion, geostatistical techniques such as kriging or co-kriging can be applied to predict the values of attributes at un-sampled sites (Ma *et al.*, 1999).

There are many different applications of geostatistics ranging from the classical fields of mining and geology to soil science, hydrology, meteorology, environmental sciences, agriculture, and also structural engineering. As time goes, the advanced built of computation facilities and the various kind of geostatistical software has making the spatial analysis of environmental popular (Nas and Berktay, 2010).

2.4.2 Kriging

In reality, there is no perfect or complete data collection at every single points for the scope area because we cannot avoid the actual constraints. Therefore, interpolation is a fundamental and essential method to do analysis, graphing and comprehension of 2 dimensional data. For the word "Kriging", it means the optimal prognostication. Kriging method will predicts unknown values from the data obtained at respective places. This method presented by semivariogram to express the spatial variation and has an advantage on minimizes the error of predicted values reckons by the spatial distribution of the predicted values (Lang, 2014).

On the other hand, Kriging method are also can be known as the best linear unbiased estimator since it will produce a result with minimum or even zero error. It has an objective that is to minimize the error variance where moving average or inverse distance weighting (IDW) can be calculated accurately in an effective way. For example, Kriging will creates weights encounter the measured values so that unmeasured locations can be determined their values. Kriging will weights for the surrounding of the measured points in a more sophisticated way than that of IDW because IDW only apply a simple algorithm based on distance while Kriging weights come from a semivariogram that was developed by viewing the spatial structure of the data (Al-Mashagbah *et al.*, 2012).

Actually, there are some spatial interpolation methods for spatial distribution of water quality parameters and they are IDW, Radial Basis, Local Polynomial, and Kriging. Kriging works the best because it can get extra information with the usage of random function. The flexibility of Kriging depends on the input parameters and it also does autocorrelation and errors of the prediction. Here are some types of Kriging for example Simple kriging, Universal Kriging, Block Kriging, Ordinary Kriging, Co-kriging and Disjunctive Kriging (BenthungoMurry, 2013).

Interpolation steps can be simple mathematical models like inverse distance weighting, trend surface analysis, Thiessen polygon and so on. For the more complex models of geostatistical methods they will be Kriging and thin plate splines. Kriging may simplify both the statistical and mathematical properties of the samples. It gives value to the spatial autocorrelation among the sample points and also together in the spatial configuration of the samples around the forecast location. Kriging is flexible and allow for the calculation of spatial autocorrelation for the data due to the use of statistical models.

The basic assumption in Kriging is about the data comes from a stationary random process and some formulae require that the data must be normally distributed. Furthermore, Kriging is divided into two different tasks that are quantifying the spatial structure of the data and producing a predicted surface model. Thus, the prediction of an unknown value for a specific point is through Kriging where they will use the best fitted model from semivariogram according to different types of model. Among all the different forms of Kriging, Ordinary Kriging has been used widely because it is known as a reliable estimation method (Nas, 2009).

Kriging method is founded and named in the honor of Dr. Kriging who was a noted South African mining geologist. With this invention, he used semivariogram to represent the estimates from a specified number of adjacent data with considering the interdependence. This application also have arisen regarding the study of spatial distributions soil water and salt in field, the interpolation to river spatial database, the spatial variance pattern of chemical substance circulation in swamp, the distribution of contaminant concentrations, and stream flow estimation. In addition, the average Kriging standard deviation use to estimate the error can be used as a measure of network effectiveness. For an optimal monitoring network, the standard deviation of error will be always the minimum. To determine the network density based on the Kriging method, there are two steps to be followed that are simulation and optimization. The benefit of the Kriging - based approach is that we can design the network as a test first without affecting the measurement of the estimation because the variance of estimation error is independent of actual measurements. However, Kriging does not consider on the physical information of